浙江大学

Fixed Income Analytics and Models：Group Project

姓名与学号 管家乐 3180103296 3180103296@zju.edu.cn

黄杰章 3180103159 1748912080@qq.com

海 日 3180101853 belieberhr@163.com

郑凯祎 3180103285 3180103285@zju.edu.cn

范咏洋 3180105911 957151769@qq.com

胥鑫凤 3180105457 3180105457@zju.edu.cn

常庭瑄 3180101879 3180101879@zju.edu.cn

朱辰邦 3180104009 462797526@qq.com

指导教师 许奇

展示日期 2020.12.10

1. Introduction

XGBoost, which means Extreme Gradient Boosting, is a scalable end-to-end tree boosting system. In our project, we use it as the main model to predict the excess return on US Treasury bonds.

In terms of data selection and processing, through a lot of literature reading, we find diverse and high-quality factors, which cover a wide variety of information, such as the investor sentiment index, unstructured data and so on. Before tuning the parameters of the XGBoost model, we perform feature engineering processing on these factors, and a validation set is added to prevent the occurrence of overfitting. Moreover, to test the performance of the predictions in the actual transactions, we use the predictions of the model in the out-of-sample/test set to construct a long-short strategy.

In the robustness checks of the model, we use the performance of OLS, Ridge, Lasso and Elastic Net in the out-of-sample/test set to demonstrate the advantages of machine learning. And also, the frequency of data, the forecasting horizons and the financial crisis is considered in this part, in which results also show the robustness of our model.

2.Literature review

For many years, researchers have been trying to increase the accuracy of predicting the bond risk premium by taking both external factors and human instincts into consideration. Fama and Bliss (1987) manage to use forward spreads to confirm the predictive power of forward rates which makes the forecast model more feasible. Cochrane and Piazzesi (2005) find the same linear combination of forward rates of different maturities to predict bond returns, extending the classic model of Fama and Bliss. Further, Ludvigson and Ng (2009) confirm that "real" and "inflation" macroeconomic factors extracted from the scope of 132 macroeconomic variables can significantly improve the predictive power of the model established by Cochrane and Piazzesi. While the three models mentioned above focus on the contribution of financial variables to bond risk premium forecast, Jeremy Goh, Fuwei Jiang, Jun Tu and Guofu Zhou (2013) attempt to introduce technical indicators used by practitioners into the forecast process which turns out that forecasting power can be effectively increased by 63 technical indicators along with economic factors. Based on Baker and Wurgler (2006), Laborda and Olmo (2014) pay more attention to the positive relationship between investors' subjective sentiment and bond risk premium, making the predictive model more creditable.

Machine learning has notable advantages in the measurement of bond risk premium. In this paper, we use the XGBoost model to develop the forecast, which implements machine learning algorithms under the Gradient Boosting framework. Apart from scalability in all scenarios, the system runs more than ten times faster than existing popular solutions on a single machine and scales to billions of examples in distributed or memory-limited settings. (Tianqi Chen and Carlos Guestrin, 2016)

3. Data and methodology

3.1 Data sources

We get U.S. Treasury bond prices from 1970.1 to 1999.12 from Fama-Bliss dataset available at the Center for Research in Securities Prices (CRSP). The data of the common macro factor and the data used to build the CP factor are downloaded from Ludvigson's NYU website. The original data of investor sentiment factor BW comes from Wugler's NYU website. And the data of the uncertainty factor is from https://www.sydneyludvigson.com/data-and-appendixes. Finally, we get a set of data about nearly 300 indicators which are detailed in https://github.com/JZ-H/ZJU_FIS_Group_Project.

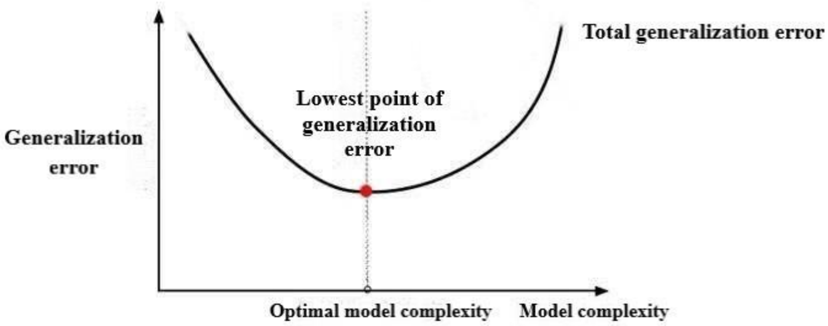
3.2 Methodology

We mainly use machine learning to predict the bond excess returns of U.S. Treasury bonds (1970-2000) and evaluates the prediction method by calculating the RMSE (Root Mean Squared Error) of the predicted value and backtesting the long-short investment strategy under the prediction method.

Specifically, first, before training the XGBoost model, we clean the original data (including nearly 300 indicators such as macro indicators, investor sentiment factor BW, uncertainty and other factors mentioned above, the frequency is monthly). Then, to improve the data quality and to make sure the training processes efficiently, we perform feature engineering processing on the factors, that is, performing PCA processing according to the correlation matrix (factors with coefficient larger than 0.8 will be compressed into one dimension, and finally obtain 73 factors) and then standardize the data. After that, we set 60% of the data (1970-1988) as the training set for training the model; 20% (1988-1994) as the validation set for parameter tuning; the remaining 20%

(1994-2000) as the test set to evaluate the prediction quality. The training set and validation set are in-sample sets, and the test set is an out-of-sample set.

To optimize the parameters of XGBoost, we use the training set and default parameters to train the model and tuning the parameters according to the RMSE calculated in the validation set. As shown in the figure below, with the complexity of the model increases, the generalization error of the model first decreases and then increases, which means the predictive ability of the model first increases and then decreases.



Therefore, we tune the parameters in order of the degree to which parameters affect the complexity of the model (the parameter with the greatest influence will be tuned first). And taking efficiency into account, we use 4 stages to get the optimal parameters. For each stage, we optimize two parameters by setting the numerical values (n, m), with which the model can gain the lowest RMSE in the validation set after training in the training set, to the two parameters.

Then we use the data in the training set and the validation set to train the model with tuned parameters, and make predictions on the test set. We measure the performance of the model in the test set based on the RMSE of the predicted value and the actual value, and thus evaluate the quality of the model.

As for the robustness checks of the model, we use the stepwise regression combined with OLS, Ridge, Lasso, and Elastic Net to predict bond excess returns. We also calculate the RMSE to compare the accuracy of prediction, between the traditional models and the machine learning.

All codes including data processing, machine learning, quantitative backtesting and so on are open source on https://github.com/JZ-H/ZJU_FIS_Group_Project (code completed independently).

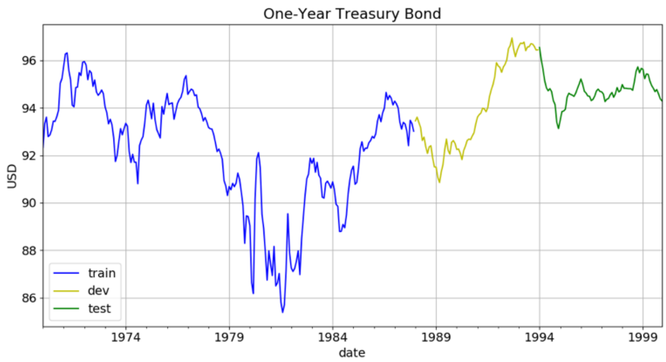
4. Empirical results and discussion

4.1 XGBoost

4.1.1 Tuning

We use the stepwise regression method to filter factors, and get CP, fs2, fs3, fs4, Finan_h1 and BW. Then we choose the XGBoost model to get the price of Treasury bonds.

Taking one-year Treasury bond as an example, we split 30 years of data into training set (60%), validation set (20%) and test set (20%). And we get the time distribution of different sets of data.

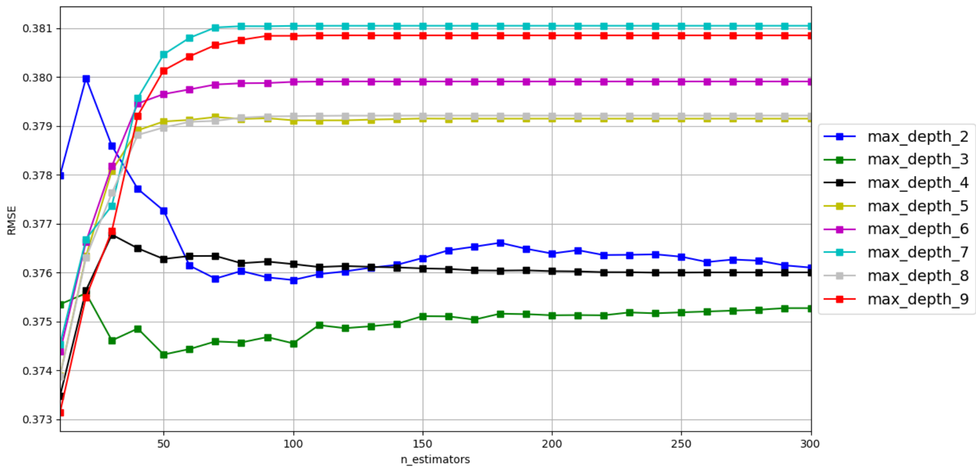


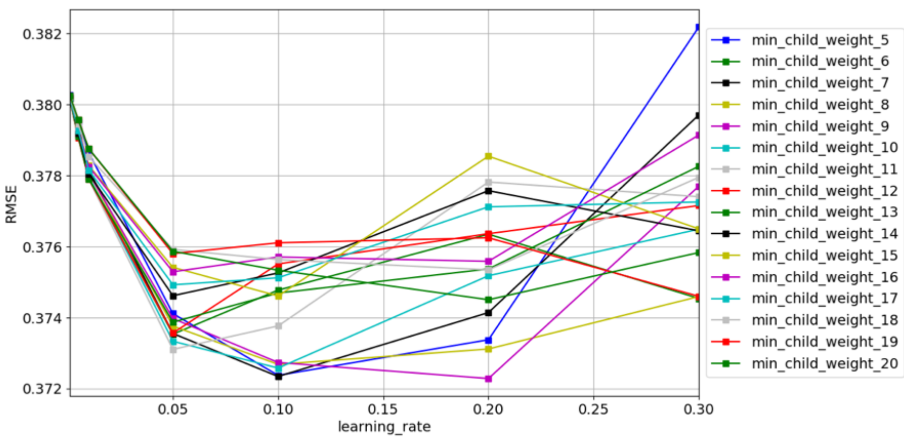
In regression problems, RMSE is a way to identify how our model would perform. To optimize the RMSE of the model, we tune the 8 parameters.[1]

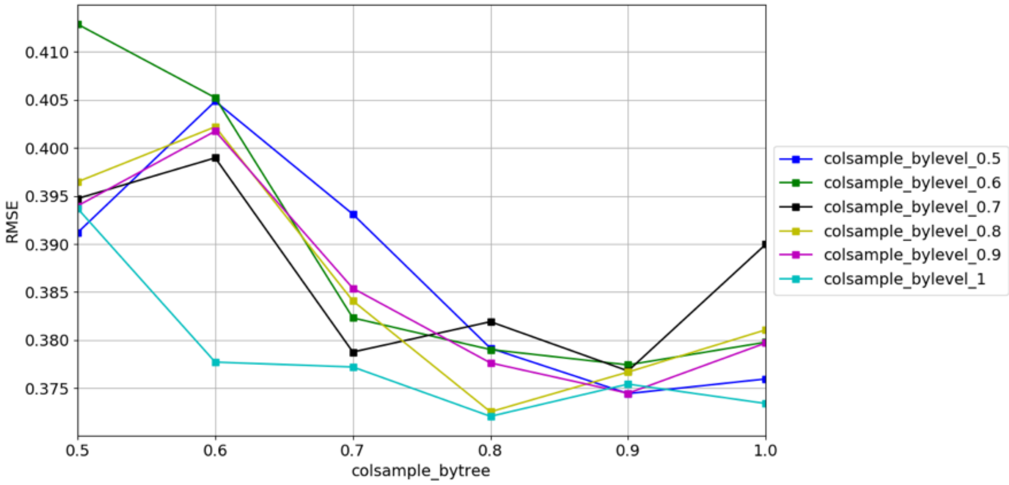
n_estimators	Number of boosted trees to fit
max_depth	Maximum tree depth for base learners.
learning_rate	Boosting learning rate (xgb's "eta")
min_child_weight	Minimum sum of instance weight(hessian) needed in a child
subsample	Subsample ratio of the training instance.
colsample_bytree	Subsample ratio of columns when constructing each tree.
colsample_bylevel	Subsample ratio of columns for each split, in each level.
gamma	Minimum loss reduction required to make a further partition on a leaf node of the tree

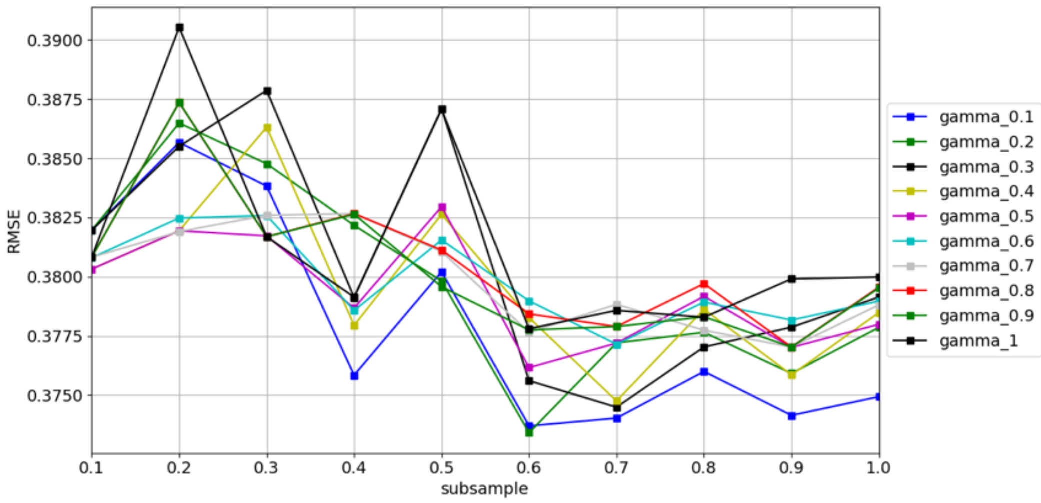
[1] The tuning is all processed on one-year Treasury bond, and others based on the remaining four assets are provided on GitHub.

First, we tune the 8 parameters in the decreasing order of importance of them. That is, n_estimators, max_depth, learning_rate, min_child_weight, subsample, colsample_bytree, colsample_bylevel and gamma. Then, considering the accuracy and efficiency of the process, the tuning is executed by passing two of the parameters. In each tuning pair, the RMSE of the model is generally V-shaped function. The optimum value of the parameter is the rock bottom of it. Figures of RMSE and parameter value are provided below (based on one-year Treasury bond).





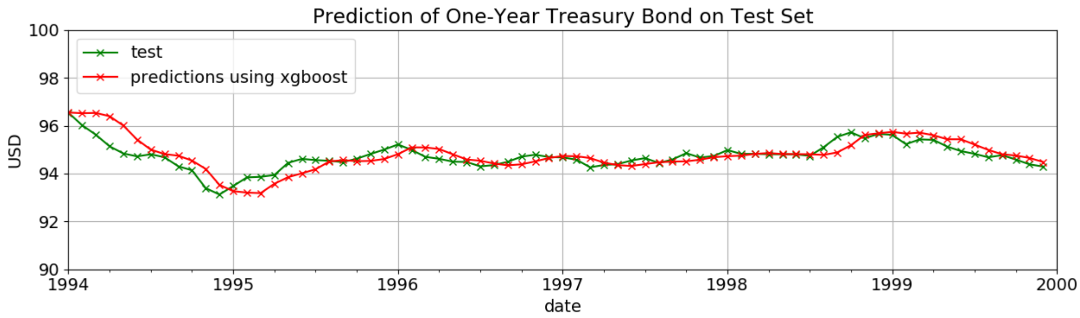


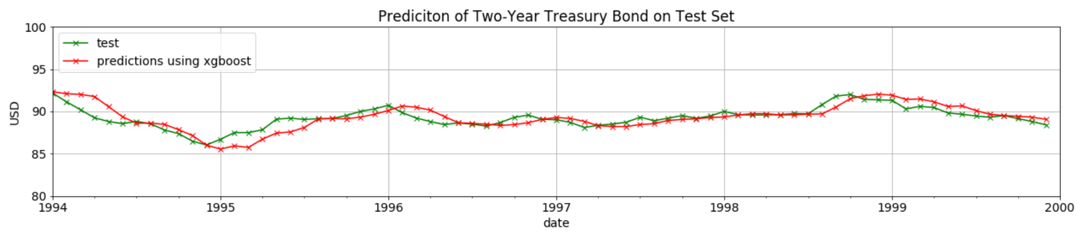


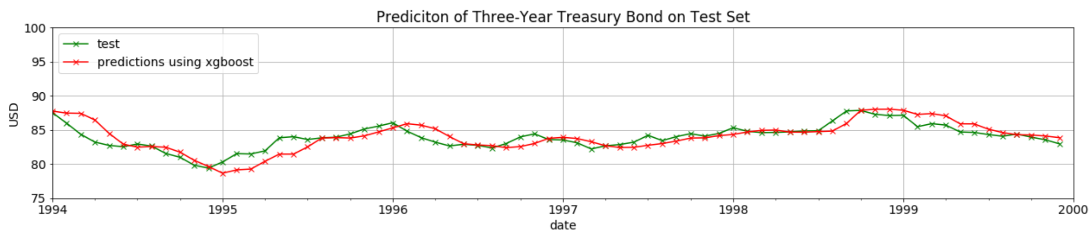
And the optimal value of parameters we get are as follows.

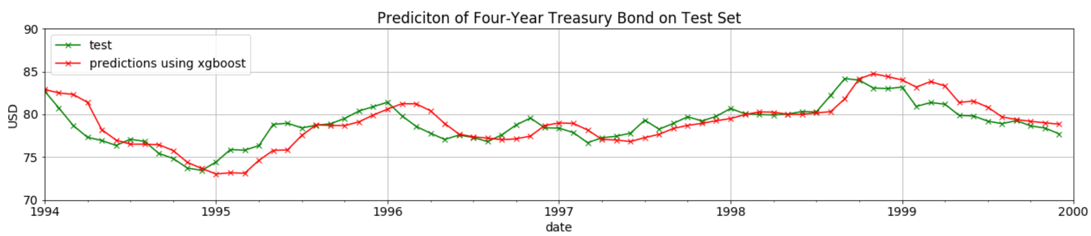
param	original	after_tuning
n_estimators	100.000	10.000
max_depth	3.000	9.000
learning_rate	0.100	0.200
min_child_weight	1.000	9.000
subsample	1.000	0.600
colsample_bytree	1.000	0.800
colsample_bylevel	1.000	1.000
gamma	0.000	0.200
RMSE	0.375	0.372

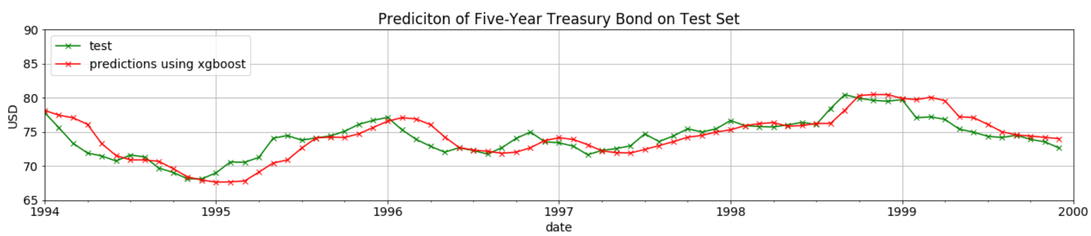
With the tuning operated successfully, the performance of the final model on the test set are:











4.1.2 RMSE

We calculate the RMSE of bond excess return to compare the accuracy of the benchmark, which is calculated according to the Expectation Hypothesis, with our model. Generally, the smaller the RMSE, the smaller the difference between the predicted and observed values. The RMSE of the benchmark is 0.031801550 on average, while that of XGBoost is 0.010584968, which means our model better fits the data.

	average	rx(2)	rx(3)	rx(4)	rx(5)
XGBoost	0.010584968	0.005863305	0.009224560	0.013061082	0.015095733
Benchmark	0.031801550	0.030112872	0.031266964	0.032629393	0.033817851

4.2 Strategy

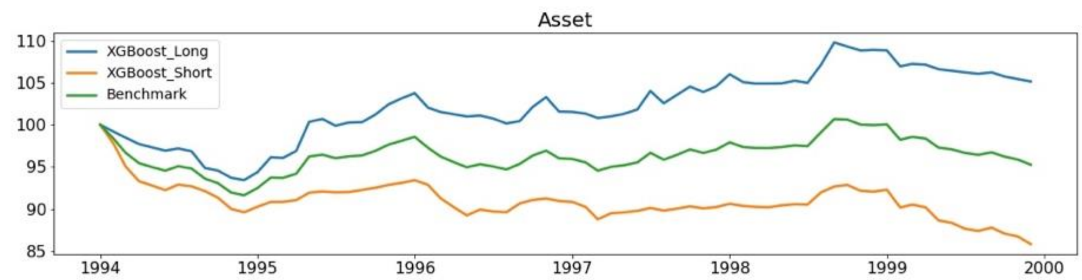
4.2.1 Long-short strategy

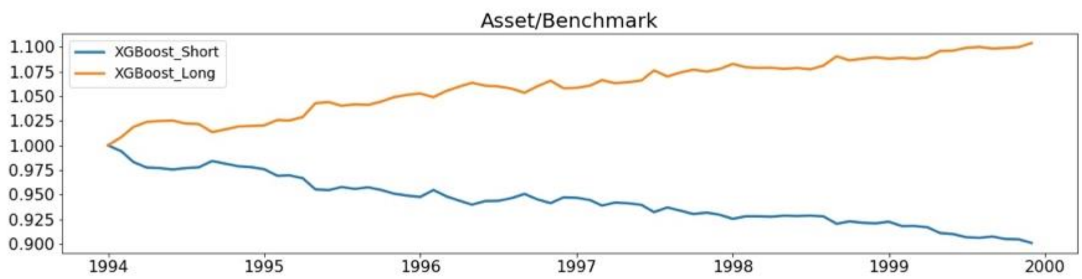
A long-short equity strategy seeks to minimize market exposure while profiting from stock gains in the long positions, along with price declines in the short positions. It works by exploiting profit opportunities in both potential upside and downside expected price moves. For all five kinds of assets, we take long positions in best two performed assets and take short positions in relatively worst two assets.

4.2.2 Backtesting for the long-short strategy

Backtesting assesses the viability of a trading strategy by using historical data to generate results and analyze risk and profitability before risking any actual capital. A well-conducted backtesting that yields positive results is able to assure traders that the strategy is fundamentally sound and is likely to yield profits when implemented in reality.

To see how our strategy would play out, the data from 1994 to 2000 are used to backtest. Meanwhile, with the equal weighted portfolio of the five treasury bonds considered as benchmark, we also backtest to compare with our long-short strategy.





For long-short strategy, we get annualized return of 0.034365, volatility of 0.033694, maximum drawdown of -0.065823 and Sharpe of 1.019916, while the results for benchmark are 0.008226, 5.752309, -0.090017 and 0.001430.

The annualized return of our long-short strategy is approximately 4 times as that of the benchmark, and the volatility is only 0.59% of benchmark's. This indicates that the long-short equity strategy can improve the return profile of the portfolio.

Moreover, the maximum drawdown and Sharpe is improved, compared with the benchmarks. A maximum drawdown refers to the maximum observed loss from a peak to a trough of a portfolio before a new peak is attained. It measures the size of the largest loss and assess the relative riskiness of one strategy versus another. For the Sharpe ratio, it is a well-reputed measure that indicates how well an equity investment performs in comparison to the rate of return on an investment. It is obvious that our long-short strategy is exposed to lower level of risk at an extreme extent and has a great improvement in the Shape ratio.

	Benchmark_average	Strategy
annualized return	0.008826	0.034365
volatility	5.752309	0.033694
max drawdown	-0.090017	-0.065823
sharpe	0.001430	1.019916

5. Robustness checks

This section provides some robustness checks to assess whether the predictability of our model in the test set is still applicable and generalized. We carry out four exercises to assess the predictability.

5.1 Regression

We compare the accuracy of prediction among the XGBoost model, the traditional model and the benchmark. We compute RMSEs of XGBoost, Benchmark, Elastic Net, Lasso, Line and Ridge models which are listed in Table 1.

Generally, the larger the RMSE, the larger the difference between the predicted and observed values, which means the worse a model fits the data. Conversely, the smaller the RMSE, the better a model is able to fit the data. The RMSEs of our regression model fluctuate across 0.01, while most of the rest are more than 0.1 and some indeed near 0.8. Even compared to the benchmark, our RMSE is much lower. Apparently, the predictive power of our model is higher than others.

Table 1.

	Root Mean Square Error				
	average	rx(2)	rx(3)	rx(4)	rx(5)
XGBoost	0.010585	0.005863	0.009225	0.013061	0.015096
Benchmark	0.031802	0.030113	0.031267	0.032629	0.033818
Elastic Net	0.243188	0.435019	0.349601	0.198017	0.732945
Lasso	0.334191	0.573986	0.557769	0.185303	0.718857
Line	0.160444	0.017421	0.016105	0.015315	0.634283
Ridge	0.160507	0.020341	0.018678	0.016274	0.634564

5.2 Data frequency

In terms of sampling frequency, we use quarterly date instead of monthly data to evaluate the bond excess returns of U.S. Treasury from January 1994 to October 1999. In this situation, owing to the expansion of the sampling frequency, the information contained in the data is reduced and incomplete, resulting in relatively poor quality. So, it’s necessary to verify the robustness in this situation of our model.

On the one hand, we compute RMSEs of this situation and compare with the RMSEs in the XGBoost in Chapter 4 and benchmark models(Expectation Hypothesis). The results are listed in Table 2 which report that the expansion of the data frequency from one month to one quarter brings about RMSEs expansion. But the difference is not sufficiently big, with an average difference of only 0.015956 which is not able to reject the hypothesis that our model using quarterly data has high predictive power.

On the other hand, we plot Figure 1 to represent the prediction of bond excess returns of five assets after changing and can intuitively see that our prediction is very close to the real data. The influence of sampling frequency change on predictive accuracy is small.

5.3 Forecasting horizons

In terms of forecasting horizons, taking the lag effect of the factors into account, we expand the forecasting horizons from one month to one quarter. From Table 2, we find that all RMSEs of this situation are smaller than the original regression model. This means that the predictive accuracy of our model increases slightly with extended forecasting horizons. At the same time, you can see this more visually in Figure 2.

5.4 Financial crisis

We consider the saving and loan crisis of 1985-1995 which lasted for a long time and cost a lost. The experience learned from the crisis played an important role in solving the subprime crisis in U.S. in 2008. Therefore, the impact of the financial crisis is also essential in testing the robustness.

From Table 2, we can also see that although the regression model is negatively affected by the financial crisis, it is better than benchmark. And the altered benchmark model performs well after the crisis. Furthermore, it's worth mentioning that in this part we split 30 years of data into training set (50%), validation set (10%) and test set (40%) which shrinks the in-sample data (to make sure the data of 1985-1995 is in the out-of-sample/test set), but our model continue to perform well. And looking at Figure 3, we can also see that our prediction curve is basically consistent with the real data. Therefore, the predictive accuracy of our model is less affected by the financial crisis.

Table 2.

	Root Mean Square Error				
	average	rx(2)	rx(3)	rx(4)	rx(5)
Situation1: sampling frequency（use quarterly data instead of monthly data）					
XGBoost	0.010585	0.005863	0.009225	0.013061	0.015096
XGBoost (altered)	0.026541	0.015459	0.023009	0.030922	0.037722
Benchmark	0.031802	0.030113	0.031267	0.032629	0.033818
Benchmark(altered)	0.040662	0.035036	0.038843	0.043115	0.047057
Situation2: forecasting horizons（one quarter ahead instead of one month ahead）					
XGBoost(altered)	0.010422	0.005838	0.009034	0.012479	0.015356
Benchmark(altered)	0.030200	0.028974	0.029797	0.030831	0.031814
Situation3: financial crisis					
XGBoost(altered)	0.01518	0.008483	0.012956	0.01774	0.022089
Benchmark(altered)	0.029916	0.028079	0.029235	0.030872	0.032665

Fig. 1. Data frequency

Fig. 2. Forecasting horizons

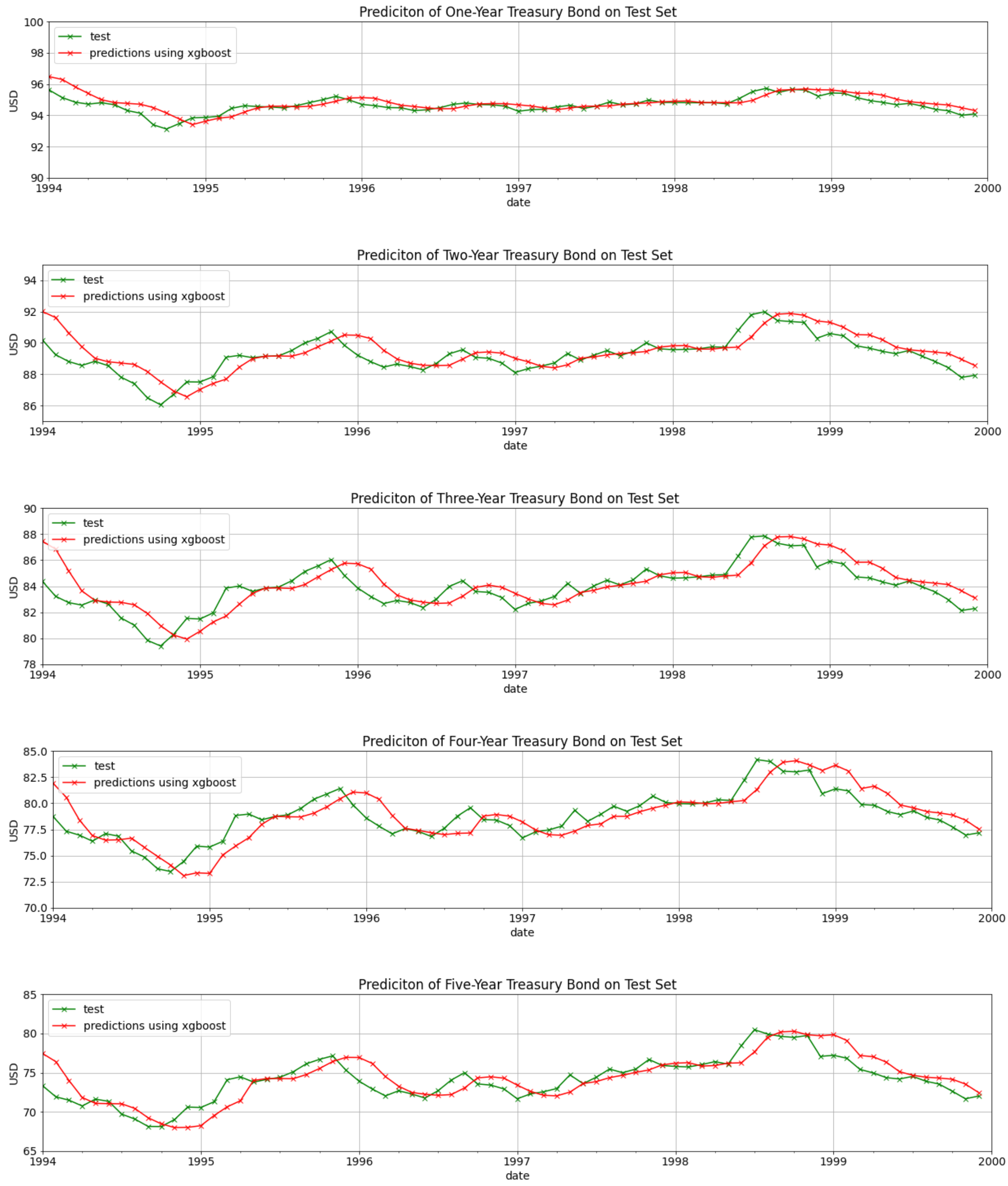
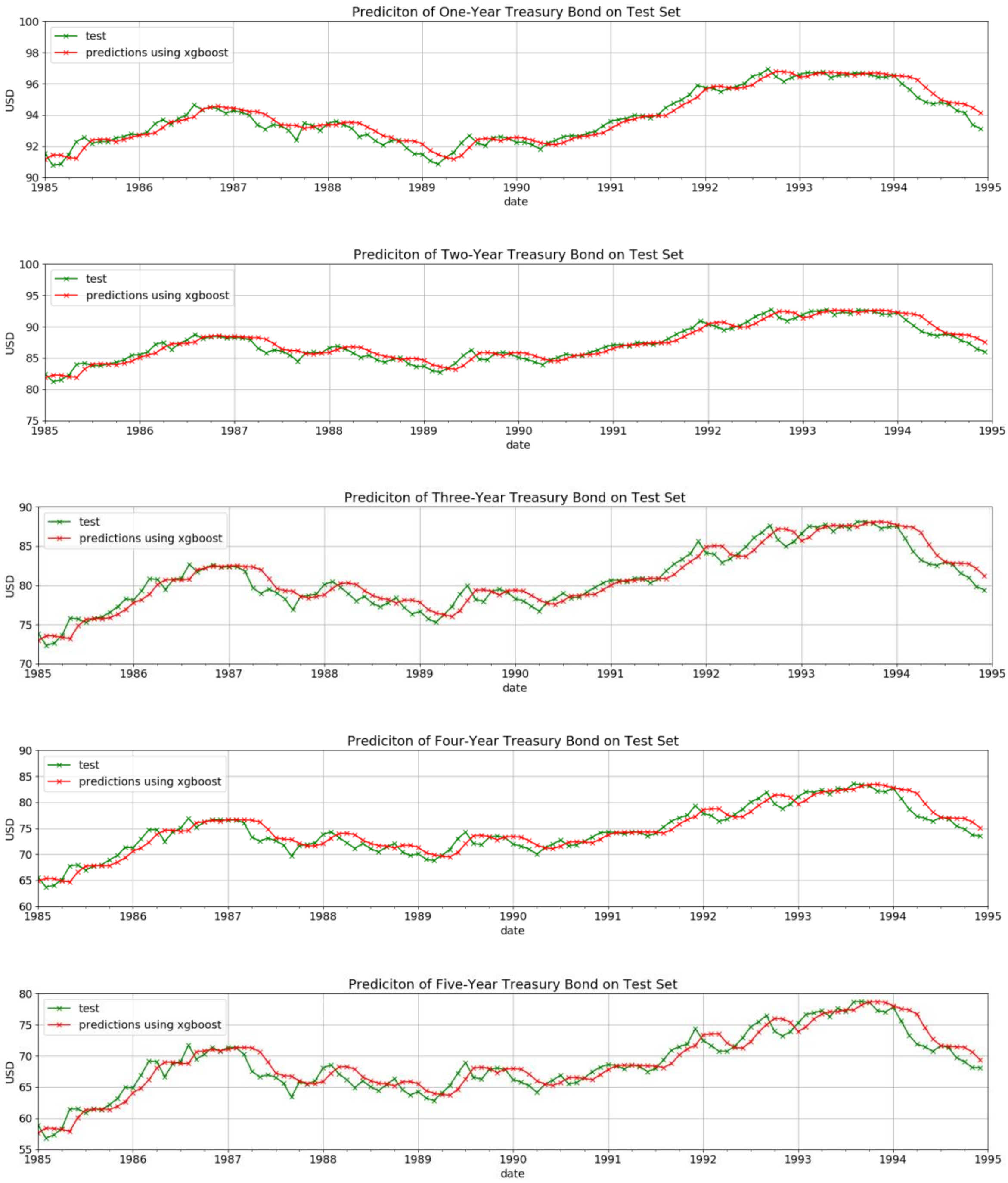


Fig. 3. Financial crisis



To sum up, our model has wonderful predictive accuracy under various conditions.

6. Conclusion

Generally speaking, we use machine learning method to forecast the excess return of U.S. Treasury Bond (1970-2000), and calculate the RMSE of the predicted value, and construct the long-short strategy.

Firstly, we calculate the RMSE of bond excess return to compare the accuracy of the benchmark and our model. And the results of RMSE in the out-of-sample/test set means our model has a better prediction effect.

Secondly, in terms of the feasibility of the strategy, we conducted a backtest with a long-short strategy. At the same time, the equal-weight portfolio of the five Treasury bonds is used as a benchmark. According to the results of the backtesting, our long-short strategy performs much better than the benchmark. This shows that the long-short strategy and the predictions of our model can effectively improve the return of the portfolio.

Thirdly, our model also performs well in the robustness checks, which is demonstrated by the RMSE of the predictions made under the change of sampling frequency, the change of forecasting horizon and the impact of the financial crisis. The RMSE shows a relative increasement only in the situation of the change of sampling frequency but still much lower than the benchmarks. Moreover, we use the stepwise regression combined with OLS, Ridge, Lasso, and Elastic Net to demonstrate the superiority of XGBoost.

When considering model deficiencies, we must admit that though our model has better predictive power than traditional models, there is a certain deficiency in the explanatory power. Our current model could not tell how exactly each factor plays a role. We hope to improve the explanatory power of this model in future research.

References

Cochrane, J. H., and M. Piazzesi. 2005. Bond Risk Premia. *American Economic Review* 95:138–60.

Fama, E. F., and R. H. Bliss. 1987. The Information in Long-Maturity Forward Rates. *American Economic Review* 77:680–92.

Goh, Jeremy and Jiang, Fuwei and Tu, Jun and Zhou, Guofu, Forecasting Government Bond Risk Premia Using Technical Indicators (July 28, 2013). 25th Australasian Finance and Banking Conference 2012, Asian Finance Association (AsFA) 2013

Conference, Available at SSRN:

https://ssrn.com/abstract=1914227 or http://dx.doi.org/10.2139/ssrn.1914227

Ludvigson, S.C., Ng, S., 2009. Macro factors in bond risk premia. *The Review of Financial Studies* 22, 5027–5067.

Ricardo Laborda, and Jose Olmo. 2014. Investor sentiment and bond risk premia. *Journal of Financial Markets* 18, 206-233.